POINT CLOUD QUALITY ASSESSMENT METRIC BASED ON ANGULAR SIMILARITY

Evangelos Alexiou and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
Emails: FirstName.LastName@epfl.ch

ABSTRACT

The rise of immersive technologies has been recently fuelled by emerging applications which employ advanced content representations. Among various alternatives, point clouds denote a promising solution which has recently drawn a significant amount of interest, as witnessed by the latest activities of standardization committees. However, subjective and objective quality assessments for this type of content still remain an open problem. In this paper, we introduce a simple yet efficient objective metric to capture perceptual degradations of a distorted point cloud. Correlation with subjective quality assessment scores carried out by human subjects shows the proposed metric to be superior to the state of the art in terms of predicting the visual quality of point clouds under realistic types of distortions, such as octree-based compression.

Index Terms— Point cloud, objective quality metrics, quality assessment

1. INTRODUCTION

In modern information technologies and communication systems, new visual modalities are employed to provide enhanced experiences to the users and faithfully reproduce real-world sceneries. Point clouds have emerged as a viable representation of immersive contents thanks to the ease in capture and processing of such content. The commercial availability of high resolution displays, along with the recent launch of low-cost depth sensors (e.g., Kinect, Intel RealSense) and their integration in hand-held devices, such as smartphones and laptops (e.g., iPhone X, HP Spectre x2), has reinforced the interest for this visual data representation.

As in any type of content, quality evaluation of point clouds is critical, as it shapes the way scenes are acquired, compressed, processed and rendered. The quality of a content is typically assessed through either subjective or objective testing. Objective quality assessment of point clouds is mainly carried out by full-reference metrics, which can be distinguished into three categories: (a) point-to-point, (b) point-to-plane, and (c) point-to-mesh [1]. The latter are sub-optimal approaches and will not be considered, as there is no unique way to reconstruct an object from a set of points, and the objective scores heavily depend on the selected algorithm for mesh generation.

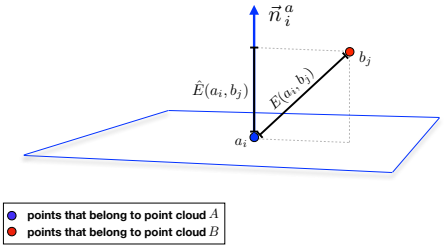


Fig. 1: Distances in the current objective metrics.

The point-to-point metrics are based on geometric distances of associated points between the reference and the content under evaluation. In particular, after identifying for each point b_j of the content under evaluation B its nearest neighbor a_i from the reference point cloud A, the Euclidean distance $E(a_i, b_j)$ is calculated indicating the error between points a_i and b_j, as illustrated in Figure 1.

The point-to-plane metrics are based on the projected error along the normal of a reference point and, essentially, larger costs are assigned to points that deviate from the underlying surface. After identifying for each point b_j of the content under evaluation B its nearest neighbor a_i from the reference point cloud A, the projected error $\hat{E}(a_i, b_j)$ is derived across the normal of the reference point, $\vec{n}_i^a$, as demonstrated in Figure 1. The point-to-plane metrics require the presence of the normal vectors of the original content. In case the latter is set as the reference, the computation is straight forward, as explained above. Otherwise, if the distorted point cloud is set as the reference, its normals are estimated by averaging over the normals of associated nearest neighborhoods that belong to the original content. Then, the projected error is calculated across this average normal vector.

In both point-to-point and point-to-plane metrics, an indi-

This work has been conducted in the framework of ImmersiaTV project under the European Union Horizon 2020 research and innovation program (grant agreement no. 688619) and funded by Swiss State Secretariat for Education, Research and Innovation SERI.

978-1-5386-1737-3/18/$31.00 ©2018 IEEE

vidual distance is associated with every point of the content under evaluation, and the geometric similarity between the test contents is expressed through an error value. This error is computed either by taking the average of the distances associated with the points of the content under evaluation (Root Mean Squared - RMS), or by taking the average squared distances (Mean Squared Error - MSE), or the Hausdorff distance[1]. Commonly, both the original and the processed stimuli are set as reference and both error values are calculated; then, the symmetric error is obtained by keeping the maximum out of these values. Finally, for objective quality assessment of contents of different scale, the Peak-to-Signal Noise Ratio (PSNR) has been proposed. In the literature, PSNR is defined as the ratio of the squared maximum distance of nearest neighbours of the original content (potentially, multiplied by a scalar), or the squared distance of the diagonal of the minimum bounding box, divided by the squared error value (i.e., squared RMS, MSE or squared Hausdorff distance).

As it was shown in [2], the current objective metrics are able to capture geometric divergences, but they fail to accurately predict the visual quality of point clouds after point removal and structural loss. In this paper we propose a new objective quality metric based on the angular similarity between two point clouds, which is falling into a fourth category, namely, plane-to-plane. Correlation with subjective scores reveals that the proposed approach outperforms the state of the art in predicting the visual quality of geometry-only contents after applying octree-based compression, while its performance remains comparable to current metrics in presence of Gaussian noise. Limitations of the proposed algorithm are reported, along with the issues to be addressed in future.

2. PROPOSED OBJECTIVE METRIC

In this section, the proposed metric is described along with complexity details, followed by its limitations.

2.1. Definition

The human brain naturally tends to interpolate points in order to perceive the underlying object. Thus, a measurement of the perceptual degradation of a distorted point cloud with respect to its original version could be obtained by estimating the similarity of the fitting surfaces that can be used to reconstruct these objects. This is comparable to geometry-based mesh-to-mesh metrics which have already been proposed in the literature (e.g., [3]). However, simply adopting mesh-to-mesh approaches to quantify the perceptual difference between two point clouds brings the issue of sensitivity to the reconstruction algorithm employed to create the corresponding meshes.

A simpler way would be to consider the tangent planes between associated points that belong to two point clouds

[1]The Hausdorff distance is defined as the maximum of the distances of each point in one set from its nearest neighbor in the other set.

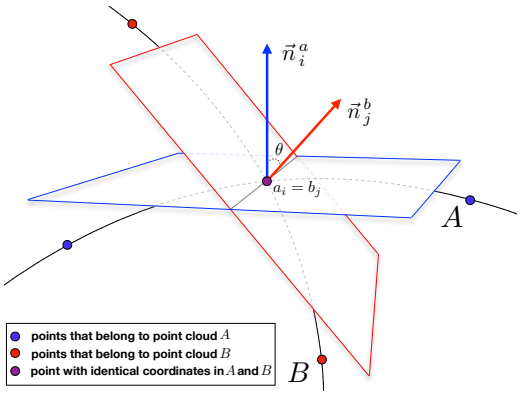


Fig. 2: Angular similarity in the proposed metric.

contents, instead of the surfaces that fit in the corresponding neighbourhoods. Tangent planes can be viewed as a local linear approximation of the underlying surface of the model, and the angular similarity between two tangent planes provides an indication of the difference between local surfaces that pass by the corresponding points. Thus, a metric that would examine the angular similarity between every pair of associated points of two point clouds, may capture the perceptual difference between the inferred 3D objects.

Let us consider a point a_i, with its associated normal vector $\vec{n}_i^a$, that belongs to a set of points A which represents an object O_A. Let us also consider another point b_j, with its normal vector $\vec{n}_j^b$, that belongs to another set of points B which represents another object O_B. Let us finally assume that the coordinates of a_i and b_j are identical, as shown in Figure 2. The difference between normals $\vec{n}_i^a$ and $\vec{n}_j^b$ is expressed through the angle θ, which is equal to the angle between the corresponding tangent planes perpendicular to these normals. Differently oriented tangent planes indicate that different local surfaces connect the points a_i and b_j with their corresponding neighbors in sets A and B. Thus, a larger angle θ implies a larger difference between the local surfaces in objects O_A and O_B, respectively.

Let us assume that point cloud A is the original content and point cloud B is a degraded version. By setting the original point cloud as the reference, for each point b_j that belongs to the distorted point cloud, the nearest point a_i of the original content is identified using the Euclidean distance. The angular similarity of a pair of associated points a_i and b_j is defined as $1 - \hat{\theta}/\pi$, where $\hat{\theta} = \arccos(\cos(\theta))$ is measured in radians and denotes the estimated angle between the tangent planes belonging to points a_i and b_j; notice that different notations are used, as $\theta \in [0, 2\,\pi]$ while $\hat{\theta} \in [0, \pi]$ by the definition of the inverse cosine. Thus, the cosine similarity $x = \cos(\theta)$ is initially derived using the normal vectors $\vec{n}_i^a$ and $\vec{n}_j^b$, based on Equation 1,

$$x = \cos(\theta) = \frac{\vec{n}_i^a \cdot \vec{n}_j^b}{||\vec{n}_i^a||\,||\vec{n}_j^b||} \tag{1}$$

with $x \in [-1, 1]$. Then, to compute the angular similarity, the inverse cosine of the obtained value x is calculated. Considering that we are only interested in the angular similarity between the tangent planes, we want to keep the minimum out of the two angles that can be formed between the intersecting planes; thus, we define $\tilde{\theta} = \min\{\hat{\theta}, \pi - \hat{\theta}\}$, with $\tilde{\theta} \in [0, \pi/2]$. An equivalent expression is given by Equation 2,

$$\tilde{\theta} = \arccos(|x|) \tag{2}$$

where x is calculated using Equation 1. Finally, the angular similarity is bounded in the range $[0, 1]$ through Equation 3.

$$\text{Angular similarity} = 1 - \frac{2\,\tilde{\theta}}{\pi} \tag{3}$$

After associating a similarity value to each point of the point cloud under evaluation, in this case B, the weighted average (WAVG) can be calculated. In an analogous way, the same computations are repeated after setting as reference the distorted point cloud. Finally, the symmetric error is obtained by keeping the minimum out of these two WAVG similarity values. The description of the metric is summarized in Algorithm 1 while a prototype implementation can be found in the following repository: `https://github.com/mmspg/point-cloud-angular-similarity-metric`.

Algorithm 1

Set as reference point cloud $A = \{a_1, a_2, ..., a_k\}$
for all j such that $1 \leq j \leq n$ **do**
    Identify a_i as the nearest neighbor of b_j in A
    Compute angular similarity $s_{b,a}(j)$ using Equation 3
Compute WAVG similarity: $\bar{s}_{B,A} = \frac{\sum_{j=1}^{n} w_j \cdot s_{b,a}(j)}{\sum_{j=1}^{n} w_j}$
Set as reference point cloud $B = \{b_1, b_2, ..., b_n\}$
for all i such that $1 \leq i \leq k$ **do**
    Identify b_j as the nearest neighbor of a_i in B
    Compute angular similarity $s_{a,b}(i)$ using Equation 3
Compute WAVG similarity $\bar{s}_{A,B} = \frac{\sum_{i=1}^{k} w_i \cdot s_{a,b}(i)}{\sum_{i=1}^{k} w_i}$
Compute symmetric error $s = \min\{\bar{s}_{B,A}, \bar{s}_{A,B}\}$

2.2. Complexity

The complexity of the proposed algorithm is limited by the selection of the algorithm to identify nearest neighbors. In particular, let us set point cloud A as the reference content. Assuming a linear search approach, the computational complexity to specify a nearest neighbor a_i for a point b_j, would be $\mathcal{O}(k)$. Following a *k*-d tree approach, a space-partitioning data structure of k points should be initially constructed, which is an operation of $\mathcal{O}(k \log k)$. Then, a search in the *k*-d tree to determine a nearest neighbor a_i for a point b_j, is an operation of $\mathcal{O}(\log k)$. After establishing a pair of associated points, the angular similarity is computed. Considering that the calculation is constant with regards to the number of points, a cost of $\mathcal{O}(1)$ is added on the top. The aforementioned procedure, excluding the potential *k*-d tree construction, is repeated $\forall\, j \in \{1, ..., n\}$. Then, the computation of the WAVG similarity poses an additional complexity of $\mathcal{O}(n)$, as it is a function of the number of points of the point cloud under evaluation B. Analogously, we obtain the computational costs after setting point cloud B as the reference. The consolidated computational complexity of the proposed algorithm is $\max\{\mathcal{O}(k \log k), \mathcal{O}(n \log n)\}$, assuming a *k*-d tree, or $\mathcal{O}(n\,k)$ assuming a linear search approach for the identification of the nearest neighbors, and is asymptotically identical to the costs of the state-of-the-art metrics.

2.3. Limitations

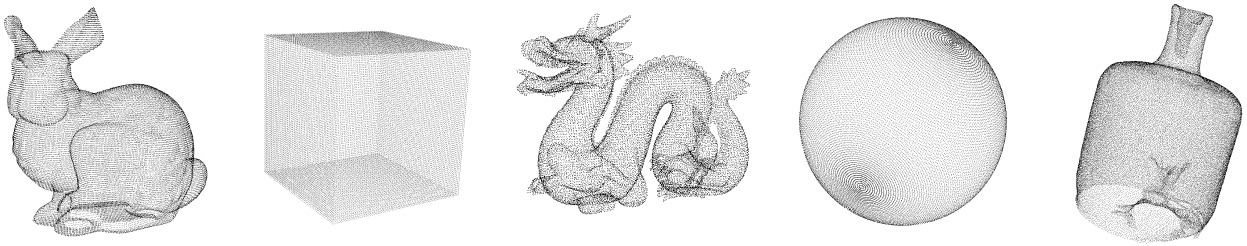
The main limitations of the proposed metric are: (a) It is a full-reference metric for geometry-only degradations, indicating that both the original and the distorted contents should be available in order to compute an objective quality score for a degraded content. (b) The angular similarity is calculated between pairs of points that are associated as nearest neighbors using the Euclidean distance, which implies that it is vulnerable to distorted point clouds by translation or scaling. (c) A normal vector should be present for every point of both the original and the distorted point cloud. This limitation indicates that in case normals do not coexist with the coordinates, they should be estimated. This dependency makes the performance to be affected by the selected normal estimation algorithm and its configuration. However, it should be emphasized that no specific methodology is imposed as part of the objective metric. Although several techniques are available, the investigation of the optimal choice is outside of the scope of this paper. Based on the results of Section 4, we show that even in the case of absence of normals, the predictive power of the proposed metric is high, after using a common and simple normal estimation algorithm with a typical configuration.

3. VALIDATION METHODOLOGY

In this section, we describe the design of a subjective quality assessment experiment to benchmark objective quality metrics. For more details, the reader may refer to [2].

3.1. Subjective experiment

A set of five simple geometry point clouds was selected to be used in experiments that were conducted in a desktop setup. The test contents were displayed as a collection of points. The models have a density in the same order of magnitude and are scaled to fit in a minimum bounding box of size 1. In particular, *cube* and *sphere* were artificially generated in order to account for synthetic contents with perfect geometry. *Vase*, was manually scanned using an Intel RealSense Camera R200 and



(a) *Bunny* (35947) (b) *Cube* (30246) (c) *Dragon* (22998) (d) *Sphere* (30135) (e) *Vase* (36022)

Fig. 3: Selected contents.

represents a typical irregular content acquired from a low-cost consumer device. Finally, *bunny* and *dragon*, were selected from the Stanford 3D Scanning Repository[2]. The geometry of these models is less irregular with very limited amount of noise and smooth underlying surfaces. In Figure 3, the selected test contents and their number of points are illustrated.

Two radically different types of degradations were selected, namely, (i) Guassian noise which is a widely used model for position errors, and (ii) octree-based compression that leads to sparser versions with structural loss. For Gaussian noise, the coordinates of every point is displaced along every dimension according to a target standard deviation $\sigma = \{0.0005, 0.002, 0.008, 0.016\}$. For octree-based compression, a suitable level-of-details is selected per content to maintain a target percentage of remaining points $p = \{30\%, 50\%, 70\%, 90\%\}$ with acceptable deviation of $\pm 2\%$. This type of distortion is annotated as octree-pruning. The degradation levels were selected in order to represent a wide range of visible distortions. In Figure 4, degradations of *bunny* are indicatively presented. The structural loss after octree-pruning can be noticed, mainly, in the upper part of the model's body, from the selected viewport.

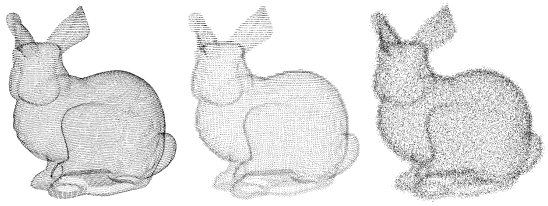


Fig. 4: *Bunny*: (left) original, (middle) octree-pruning with $p = 50\%$, (right) Gaussian noise with $\sigma = 0.008$.

The experiments were conducted in a laboratory that fulfills the ITU-R Recommendation BT.500-13 [4] for subjective evaluation of visual data. A visualizer based on VTK library and integrated in PCL [5] was used, along with an Apple Cinema Display of 30 inch (res. 2560x1600). The subjects were able to visualize the point clouds on the flat screen, interact using the mouse, and provide their scores using the keyboard.

[2]http://graphics.stanford.edu/data/3Dscanrep/

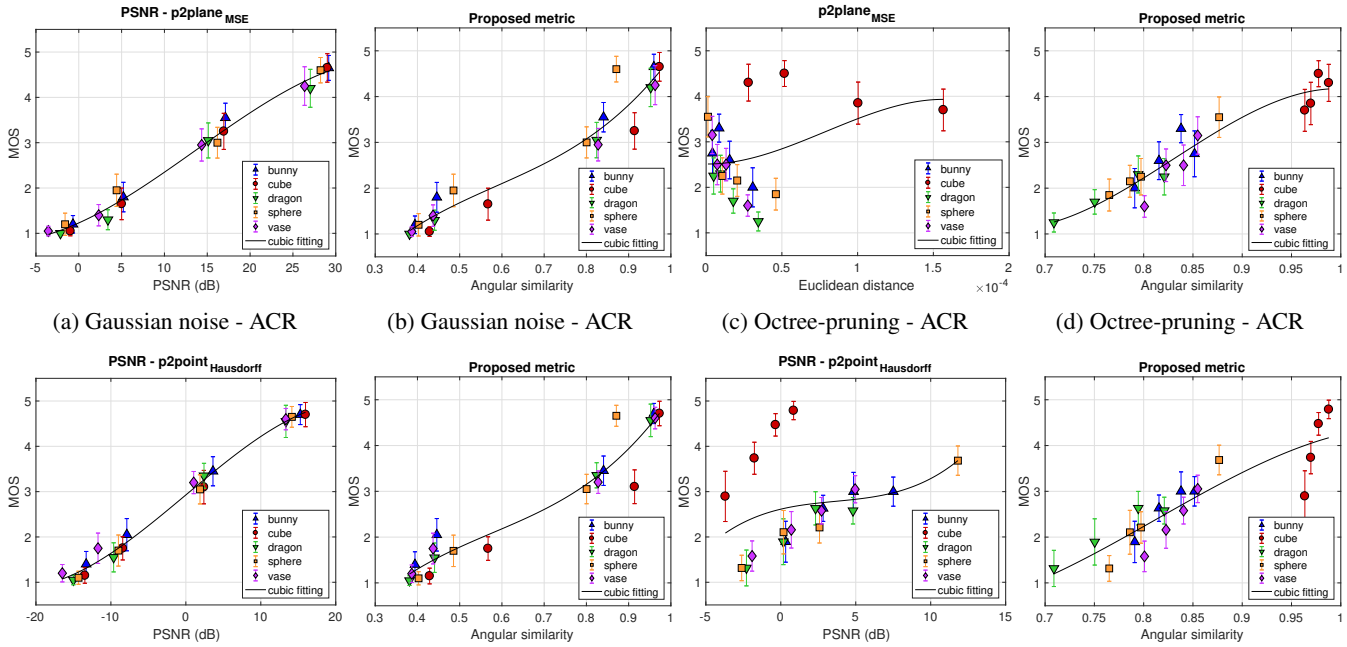
Two different subjective evaluation methodologies were selected: (i) simultaneous Double Stimulus Impairment Scale (DSIS), and (ii) Absolute Category Rating (ACR), both with 5-rating impairment scale. The first method is preferred for its high discriminative power, as the subjects simultaneously visualize both the degraded and the reference content whose position is specified. The second method accounts for a more realistic type of media consumption. As the nature of artifacts introduced by the two different types of degradations drastically differs, the experiment was split in 4 sessions.

A training phase took place before each session, allowing subjects to familiarize themselves with the the visible distortions and the evaluation tool. To avoid contextual effects and monitor defects that could introduce biases, a different permutation of the order of contents was deployed per session, the same content was never displayed consecutively, and the side of the reference in the screen was selected randomly per subject for the DSIS methodology. In each session a total of 5 contents and 4 degradation values were used, along with a hidden reference for sanity check, leading to 25 stimuli per session. A total of 20 naïve subjects participated per session, with the age ranging from 21 to 37 years old (average 28).

The subjective scores were processed by first detecting and removing outliers based on the ITU-R Recommendation BT.500-13 [4]. In DSIS with octree-pruning, one outlier was found resulting in 19 scores, while for the remaining sessions no outliers were identified leading to 20 scores. Then, the mean opinion scores (MOS) and the 95% confidence intervals, assuming a Student's t-distribution, were computed for each content under evaluation in every experiment.

3.2. Computation of the objective scores

Both the original and the distorted contents used in this experiment had no associated normals with their coordinates. To compute the normals, the methodology proposed by Hoppe et al. [6] as implemented in PCL [5] was selected. Around each point of interest 6 nearest neighbors were defined, as the test contents are rather sparse. After associating the resulting normal vectors, the obtained contents were used to calculate the scores from every objective metric. For the state-of-the-art



(a) Gaussian noise - ACR (b) Gaussian noise - ACR (c) Octree-pruning - ACR (d) Octree-pruning - ACR

(e) Gaussian noise - DSIS (f) Gaussian noise - DSIS (g) Octree-pruning - DSIS (h) Octree-pruning - DSIS

Fig. 5: Subjective against objective scores for the best state-of-the-art (a-c-e-g) and the proposed objective metric (b-d-f-h).

Table 1: Performance indexes of the best state-of-the-art and the proposed objective metric.

		Gaussian noise				Octree-pruning			
		PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
ACR	PSNR - p2plane$_{MSE}$	**0.9950**	**0.9846**	**0.1372**	**0.1000**	0.2146	0.1332	0.8933	0.6500
ACR	p2plane$_{MSE}$	0.9360	0.9379	0.4836	0.7000	0.4028	-0.0211	0.8371	0.5500
ACR	Proposed metric	0.9645	0.9466	0.3660	0.3500	**0.9530**	**0.9353**	**0.2771**	**0.1500**
DSIS	PSNR - p2point$_{Hausdorff}$	**0.9960**	**0.9782**	**0.1231**	**0.0500**	0.3364	0.4384	0.8960	0.6000
DSIS	Proposed metric	0.9676	0.9665	0.3468	0.2500	**0.8924**	**0.9032**	**0.4292**	**0.3500**

techniques, the software described in [7] was used (ver. 0.02 and 0.09). The point-to-point and point-to-plane metrics were employed, each adopting the RMS, the MSE and the Hausdorff distances. The PSNR was computed for every combination as the ratio of the squared maximum distance of nearest neighbours of the original content divided by the squared error value, leading to a total of 12 metrics. For the proposed metric, the WAVG angular similarity was calculated by setting weights equal to 1 in steps 5 and 10 of Algorithm 1.

3.3. Benchmarking of objective quality metrics

To evaluate how well an objective metric is able to estimate perceptual quality, the MOS of the participating subjects is considered as the ground truth and is compared to predicted MOS values obtained from the objective metrics. To compute the predicted MOS scores, a regression analysis using linear, logistic and cubic functions was issued, with the former providing better fitting results. Then, based on the Recommendation ITU-T P.1401 [8], the Pearson linear correlation coefficient (PCC), the Spearman rank order correlation coefficient (SROCC), the root-mean-square error (RMSE), and the outlier ratio based on standard error (OR) are computed between the subjective and predicted MOS values, to account for linearity, monotonicity, accuracy and consistency, respectively.

4. RESULTS

In Figure 5, we demonstrate scatter plots of the subjective against the objective scores as calculated from the proposed and the best-performing alternative metrics from state of the art, for each type of degradation and subjective methodology. The corresponding performance indexes are reported in Table 1. Regarding the best-performing state of the art, in the ACR experiment, the PSNR of point-to-plane with MSE, and the point-to-plane with MSE outperformed the others for Gaussian noise and octree-pruning, respectively. In the DSIS experiment, the PSNR of point-to-point with Hausdorff

distance was the best for both types of degradations. Note that point-to-point and point-to-plane metrics are annotated as p2point and p2plane, respectively, in Figure 5 and Table 1.

Based on the performance indexes of Table 1, the proposed metric outperforms the current methods in predicting the subjective visual quality under compression-like artifacts. Octree-based compression leads to elimination of high frequency components, and structural loss from point removal and displacement can be noticed. Thus, more severe visual distortions are perceived in point clouds with high curvature values and irregular structures, whereas the visual quality of regular contents with planar underlying surfaces, such as *cube*, is not significantly impacted. Although all considered metrics are able to capture visual degradations of non-planar surfaces, in case of planar contents the current metrics fail to accurately predict perceptual quality. In particular, point-to-point metrics do not consider underlying surface properties, assigning the same error value to a deviation of a point from the original position, independently of the underlying shape. Point-to-plane metrics assign different errors based on the direction of displacement of a point; that is, if a point deviates along the tangent plane perpendicular to the reference normal vector, no error occurs. However, a shift of every distorted point that belongs to a planar surface (e.g., face of a *cube*) will notably affect the objective score as a function of the shifting direction, given that same pairs of nearest neighbors are obtained, although minor visual degradations are perceived. Conversely, such errors will not significantly affect the proposed objective metric, which explains its high performance.

The high predictive power of our approach is remarkably achieved in a rather diverse point cloud data set, which consists of simple objects with regular geometry and planar surfaces (i.e., *cube*), more complex models with less regular placement (i.e., *dragon*), and point clouds with totally irregular structure (i.e., *vase*). As can be observed, better correlation results are obtained using the subjective scores collected in the ACR experiment. Note that in this methodology, the visual quality of the test contents is explicitly assessed. In the DSIS methodology, based on our observations, subjects tended to rate based on relative differences between contents, including the number of points [2].

In the presence of Gaussian noise, strong correlation between objective and subjective scores can be observed, for both the current and the proposed objective metrics. Provided that the state of the art is based on geometric distances of closest points between the reference and the distorted stimuli, by increasing the standard deviation of the noise the objective scores naturally worsen. The subjects were able to recognize such distortions and identify the level of displacement. Notably, the proposed metric achieves comparable performance, albeit the displacement of points typically leads to lower quality of normals. It should be mentioned that this high accuracy is observed, partly, as a result of the selected normal estimation algorithm, which is considered as robust against noise.

5. CONCLUSIONS

In this paper a promising alternative for objective quality assessment of point clouds is introduced, based on the angular similarity of associated points belonging to a reference and a point cloud under evaluation. Benchmarking results show that the proposed metric achieves high performance and accurately captures the visual quality of point clouds degraded by either Gaussian noise, or compression-like artifacts. For the latter, this is the only metric with high predictive power. Further experimentation on a larger set of contents and use cases is needed in order to assess its limitations. Improvements may be observed by optimizing the estimation of the normals, or by associating larger weights on areas of the point cloud that have higher impact on the perceptual quality.

6. REFERENCES

[1] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 3460–3464.

[2] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL*. International Society for Optics and Photonics, Sep. 2017, vol. 10396, p. 103961H.

[3] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "MESH: Measuring errors between surfaces using the Hausdorff distance," in *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*. IEEE, Aug. 2002, vol. 1, pp. 705–708.

[4] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan. 2012.

[5] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1–4.

[6] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface Reconstruction from Unorganized Points," in *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*. Jul. 1992, SIGGRAPH '92, pp. 71–78, ACM.

[7] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation metrics for point cloud compression," ISO/IEC JTC m74008, Jan. 2017.

[8] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul. 2012.